The Internet is a big place. It may not be as large as the universe, 78 billion light years, but it is quite probably the largest structure any of us will encounter.

I sometimes hear people compare trying to find a piece of information on the Internet to searching for a needle in a haystack. I do not find that a good analogy because searching the Internet for a "needle" will retrieve links to: the Space Needle in Seattle, Wash., needles for record players, cross-stitch supplies and medical supplies, among others.

As this example illustrates, we cannot expect to search the Internet using single, simple terms and expect to get anything resembling focused, relevant results.

In a haystack, you can use a metal detector to find your needle providing the needle contains ferrous metals in sufficient quantity and is the only metal object in the hay. The object of an Internet search engine is to transform the search back to the haystack variety, with the search engine sifting through the stack to bring back the one point of information we want. While defining an effective Internet search is more involved than waving a metal detector over a pile of hay and waiting for a beep, with the right combination of search terms it can be fairly simple.

Before we let the searching begin, I have two points of information. First, when I reference search terms, I will put them between square brackets [xx] instead of quotes. Quotation marks have a specific function in searching, which we will look at later in the article. Second, I rarely, if ever, look past the first page of search results. My experience is that relevance drops off very quickly after the first 10 results, and the links referenced on pages two through whatever tend to be redundant.

A Hunting We Will Go

For some reason, my family, friends and co-workers like to play "Stump Dale" every week or so with some obscure Internet search. For example, just before bow hunting season Zippy was wrestling with a vexing hunting dilemma: Should he use broad head or mechanical heads on his arrows?

While Zippy is an avid hunter who always gets his deer, I can barely string a bow, and my best chance of bagging a deer would be accidentally hitting one with my car. I knew what a broad head was but had not heard of mechanical arrowheads before he showed up on my doorstep with that, "You're my hero, please save me" look on his face.

He did try one search on his own, typing [which type of arrow is better] into his search engine window. What he got was apparently confusing and he wanted what he considers professional help.

So, I found myself faced with searching for information on a subject with which I have no personal experience. I find this typical of many Internet searches because we often search for information we do not already know. The trick with searching the Internet is to remember that index searching is not a technical exercise, but a semantic one.

Semantics is a discipline within the field of linguistics devoted to the study of the meaning of words, phrases and sentences. Definitions are to semantics as squares are to cubes. For example, soccer ball, ball four and Cinderella went to the ball all have different meanings for the word ball that depend on its context in relation to the surrounding words.

Zippy had the same problem with the word "arrow" because it may mean something shot from a bow or keys on a computer keyboard. If you want a relevant response from an Internet search engine, context is everything.

Let's look at the thought process behind constructing a semantically and contextually relevant search for a comparison between broad head and mechanical arrowheads for bow hunting. If you want to pull up your Web browser, open Google (the search engine I used) and follow along. Given the lag time between when I write this article — and when you read it — your results may vary.

First, let's isolate the key terms: compare, broad head, mechanical, hunting. If you put these four terms into Google and search, most of the links that come back on the first two pages will be for various commercial sites selling hunting equipment, I also found a 1999 survey of North Dakota deer hunters, but we can do better.

For the second search, we will go outside our four key terms and try the following: [broad head versus mechanical]. This returns some fairly relevant results because these terms establish a context within which the search engine can work. Google is also smart enough to find sites that use "vs." instead of versus and the term "broad head" is specific enough to archery to get good results.

There is one last refinement we can add: the word "better." Add that fourth term to the search and the results become slightly more specific because now the search engine is looking for comparisons that come to a conclusion or at least talk about relative merits.

You can also try searching on [broad head mechanical better] and get many of the same results as our previous three-word search, but you will get more advertisements for hunting gear because better shows up in advertising more than versus. Choice of terms can make a difference.

However, despite solving Zippy's search conundrum, what we found did not conclusively answer his question about which arrowhead to use. The experts who posted their test results and opinions still basically left the final decision up to the reader. The lesson here is that even though you may learn how to search for relevant
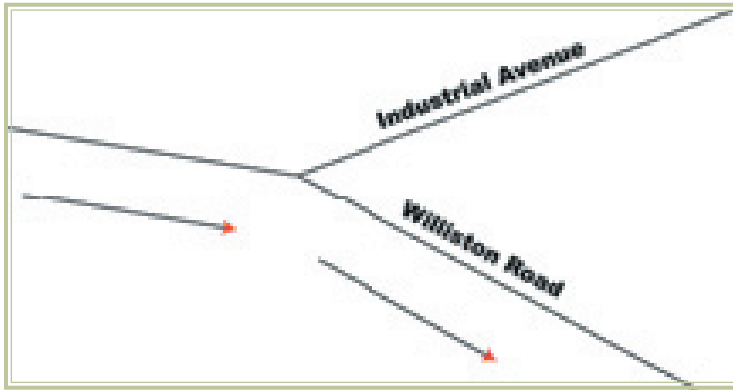
Figure 1.

results you should still be prepared to critically examine what you find and make reasoned decisions. Information, no matter how detailed, does not become knowledge until you apply it and assess the results.

Phraseology

To give you an idea just how specific information on the Internet can be, search for the right-turn-on-red rule for eastbound traffic at the intersection of Williston Road (also part of U.S. Route 2) and Industrial Avenue in South Burlington, Vt. There is a traffic light at the Industrial Avenue intersection which forms a Y junction (see Figure 1).

When the light at this intersection is green for eastbound Route 2 traffic, it is red for westbound, which means the right-of-way belongs to Industrial Avenue, not the highway. So the basic question is: Can you make a right-turn-on-red at this or any other Y-junction? Internet searching does not get much more specific than that.

First, we will try the obvious: [right turn red Y intersection], but without much luck. The second result included rules for T-junctions, with no discussion of Y. The rest of the links on the first page were directions of some type that included references to Y intersections.

Is there a way to get the search engine to filter out those results? Yes, there is. The first advanced technique we will look at is entering a search that looks for all our terms minus pages that include the word directions. It looks like this: [right turn red Y intersection -directions]. Putting a minus sign in front of a word in a search engine will tell it not to include results that include that word. Now page one still includes some links to directions, but it also now includes a link to a *Burlington Free Press* article about that intersection.

However, there is a faster way to get there. If I search on [right turn red williston road industrial avenue] the search results now return that same article at the top of the page.

The lesson here is that you will get more accurate results from specific queries than you will from generic ones. Sometimes it takes three words; sometimes it takes eight or nine. But, as with our first example, it is all about finding the right combination of terms in the correct context.

There are a host of other operators that you can use for complex searches, but the only two I have found much use for are the minus sign we used in the last example and quotation marks.

Putting quotes around a phrase tells a search engine, at least those using Boolean logic, to search for the enclosed words as a group and in a specific order.

Quotation marks are useful if you are looking for a specific document that contains a specific piece of text. If you type ["rats' feet over broken glass"] into Google as a search most of the links should take you to something about T.S. Elliot's 1925 poem, "The Hollow Men." The results without the quotation marks are not nearly as accurate.

The phrase search technique also works fairly well for finding song lyrics, speeches that contain famous quotes, and college professors trying to find if a student has simply copied something from the Internet for a paper without properly citing it. Sometimes, though, using quotes does not actually improve the search results. When I tried the same search with and without quotes in Yahoo's search engine, the results without the quotes actually were better. But because technology evolves almost daily in this area your mileage may vary.

There are a variety of other types of advanced search filters. Using Google as an example, you can filter your search results based on:

• All of the words
• An exact phrase using quotation marks
• Finding at least one of the words
• Excluding pages with particular words using a minus sign
• The language used – any of 35 different languages
• File format – any format, such as .pdf, .doc, .xls, .rtf …
• A Web page's latest update
• Numeric range – numbers between _ and _
• Where terms occur – title of the page, text, URL, links
• Domain

However, my experience has been that other than the rare need to exclude a term or find a phrase I have never needed any of the filters listed above to find what I was looking for. They are amazing displays of implementing technology — but for most of us — they are just the search engine developer showing off.

Closing Words

The search methods we reviewed should work as well for finding information on radio frequency management and federal procurement policy as they did for arrowheads and traffic regulations. Just remember the three basic concepts: understand the context of your search terms, be specific, and keep it simple. We should incorporate this technology into our own Web sites.

Knowledge may be power, but only if you can find the knowledge that you need. What Google and Yahoo can do with 20 billion pages, we should be able to do with our information repositories too.

Until next time, Happy Networking!

Long is a retired Air Force communications officer who has written regularly for CHIPS since 1993. He holds a Master of Science degree in Information Resource Management from the Air Force Institute of Technology. He is currently serving as a telecommunications manager in the U.S. Department of Homeland Security.

CHIPS